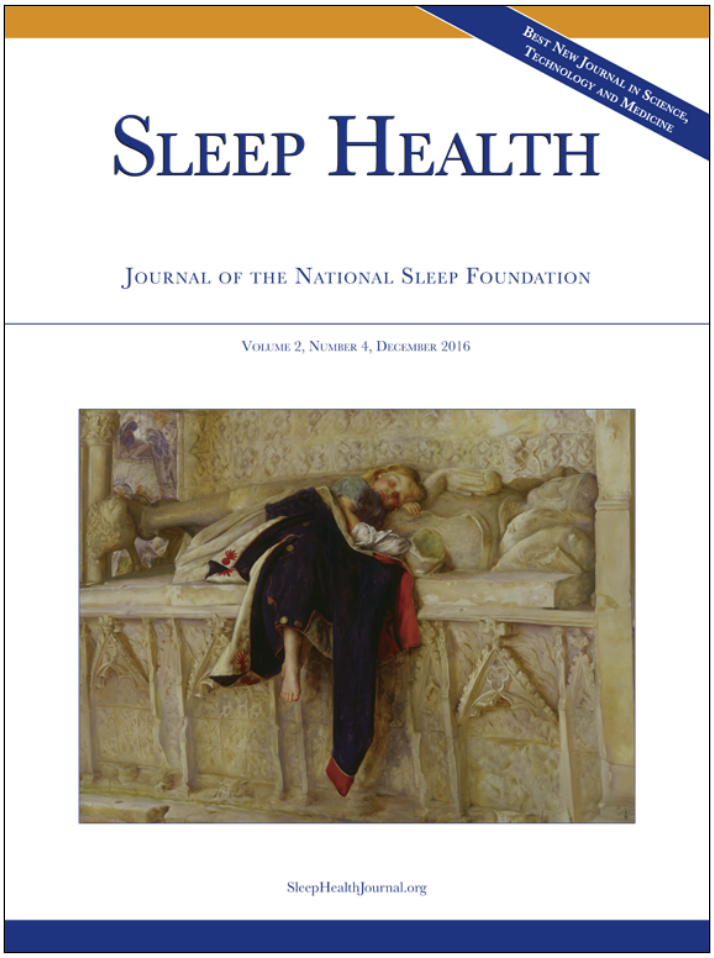

Sleep Health 2 (2016) 341–347

Contents lists available at ScienceDirect

Sleep Health

Journal of the National Sleep Foundation

journal homepage: sleephealthjournal.org

What is segmented sleep? Actigraphy field validation for daytime sleep and nighttime wake☆

David R. Samson, PhD [a,*], Gandhi M. Yetish, PhD [b], Alyssa N. Crittenden, PhD [c], Ibrahim A. Mabulla, BA [d], Audax Z.P. Mabulla, PhD [d], Charles L. Nunn, PhD [a,e]

[a] Department of Evolutionary Anthropology, Duke University, Durham, NC
[b] Department of Anthropology, University of New Mexico, Albuquerque, NM
[c] Department of Anthropology, University of Nevada, Las Vegas, NV
[d] Department of Archaeology and Heritage, University of Dar es Salaam, Dar es Salaam, Tanzania
[e] Duke Global Health Institute, Duke University, Durham, NC

ARTICLE INFO

Article history:
Received 23 June 2016
Received in revised form 10 September 2016
Accepted 26 September 2016

Keywords:
Actigraphy
Napping
Validation
Hunter-gatherer
Segmented sleep

ABSTRACT

Objective: To compare different scoring parameter settings in actigraphy software for inferring sleep and wake bouts for validating analytical techniques outside of laboratory environments.
Design: To identify parameter settings that best identify napping during periods of wakefulness, we analyzed 137 days on which participants reported daytime napping, as compared with a random subset of 30 days when no naps were reported. To identify settings that identify periods of wakefulness during sleep, we used data from a subsample of women who reported discrete wake bouts while nursing at night.
Setting: Equatorial Tanzania in January to February 2016.
Participants: The Hadza—a non-industrial foraging population.
Measurements: Thirty-three subjects participated in the study for 393 observation days. Using the Bland-Altman technique to determine concordance, we analyzed reported events of daytime napping and nighttime wake bouts.
Results: Only 1 parameter setting could reliably detect reported naps (15-minute nap length, ≤50 counts). Moreover, of the 6 tested parameter settings to detect wake bouts, the setting where the sleep-wake algorithm was parameterized to detect 20 consecutive minutes throughout the designated sleep period did not overestimate or underestimate wake bouts, had the lowest mean difference, and did not significantly differ from reported wake-bout events.
Conclusion: We propose operational definitions for multiple dimensions of segmented sleep and conclude that actigraphy is an effective method for detecting segmented sleep in future cross-site comparative research. The implications of such work are far reaching, as sleep research in preindustrial and developing societies is documenting natural sleep-wake patterns in previously inaccessible environments.

Introduction

Actigraphy, obtained via wrist-worn portable devices, is a valuable approach to investigate sleep outside of clinical or laboratory settings. Although polysomnography (PSG) is currently considered the gold standard for quantifying sleep, it remains cumbersome, expensive, and difficult to apply with ambulatory participants. Thus, actigraphy is increasingly used to study the effects of psychological disorders, stress, and disease on sleep,[1–5] and to investigate sleep in various populations living in postindustrialized Western societies.[6–8] More recently, actigraphy has been used to measure sleep in developing countries,[9,10] including studies that have characterized sleep in preindustrial societies[11,12] and high-latitude populations with long photoperiods.[13,14] Given that actigraphy-based research is a growing domain, validating accelerometry-based data in a variety of study designs has been identified as a critical goal for current sleep research.[15]

☆ Author contribution: DRS and ANC designed research; DRS and IAM performed research; DRS and GMY analyzed data; DRS, GMY, ANC, and CLN wrote the manuscript. AZPM interfaced with the study population, including translation and administration of instructions and surveys.

* Corresponding author at: Biological Sciences 107, Duke University, Durham, NC 27708. Tel.: +1 812 266 0234.
E-mail address: david.samson@duke.edu (D.R. Samson).

http://dx.doi.org/10.1016/j.sleh.2016.09.006
2352-7218/

Unlike PSG, actigraphy cannot distinguish between sleep stages, including non–rapid eye movement and rapid eye movement. However, a comprehensive literature demonstrates that actigraphy can accurately detect sleep vs wake states when compared with PSG.[6,16,17] These results clearly show that actigraphy is an efficient and cost-effective technique to study sleep in natural settings. However, most of these studies pertain to nighttime sleep and function best in the cases of monophasic, consolidated sleep. A critical question in the study of sleep in traditional societies (rural agriculturalists, pastoralists, and hunter-gatherers) concerns the operational definition of segmented sleep (ie, both daytime naps and nighttime wake bouts), and specifically how data generated by actigraphy can be used to accurately identify daytime napping and nighttime waking. It remains unclear whether actigraphy can reliably infer patterns of shorter periods of segmented sleep-wake phases (either daytime napping or nighttime waking).

Kanady et al[18] used PSG to confirm that actigraphy can be used to differentiate sleep from wake in nap and no-nap periods reliably in a clinical setting, using the "minor rest interval" feature of the Respironics Actiwatch-64 software. Other softwares have a similar feature, whereby sleep segmentation can be identified by manually setting the number of major sleep intervals per 24-hour period. However, if the software is configured to identify 2 sleep periods, it may identify those 2 sleep periods during the night, or it may infer a single bout of night sleep plus a daytime nap. The drawbacks to this process are that (1) it requires *a priori* knowledge of the presence/absence of segmentation and (2) it is nonspecific with regard to identifying daytime naps vs nighttime wake bouts. It is especially important to not require specific *a priori* knowledge of whether or not such segmentation exists when attempting to impartially investigate whether or not such segmentation occurs in field settings.

In general, actigraphy is reliable at detecting naps, but less reliable at detecting the absence of naps.[18] The most common actigraphic measures for arousal during nighttime sleep are wake after sleep onset, sleep efficiency (actual sleep time expressed as a percentage of time in bed), and sleep fragmentation (sum of mobile time and immobile bouts as a percentage). However, these measures fail to identify long, contiguous bouts of activity that are relevant to discriminating between brief arousals (a common occurrence while transitioning between sleep stages,[19] and substantial periods of nighttime awake activity where does the parenthesis close?.

Few validation studies have been performed on actigraphy and napping and, to our knowledge, no such study has been performed in nonlaboratory, field environments or across nonindustrial societies. Indeed, a previous study called for additional validation research in a wide variety of populations.[18] Circadian rhythms are stronger in non-Western populations that have more exposure to natural light and temperature cycles, and thus greater entrainment to circadian cues.[10] It is essential to validate studies in these environments, because the circadian rhythm and sleep duration among populations living in the postindustrialized West are attenuated by highly insulated, temperature- and humidity-controlled buildings in ways that may impact sleep duration and phasing. To reliably identify similarities and differences in sleep patterns across populations, it is critical that methodological issues are not producing artificial differences in sleep recordings that do not actually reflect real differences in sleep patterns in such dramatically different environments.

Several studies have used actigraphy to report daytime naps, yet operational definitions of daytime napping or nighttime waking remain elusive. Thorpy[20] defined a nap as any short sleep episode out of bed, without criteria on duration. Yoon et al[8] used this definition to report the percentage of participants that showed any napping behavior throughout their study period. Yetish et al[12] similarly reported the percentage of days in which participants napped, based on both automatic algorithmic detection of nap periods greater than 15 minutes and an additional manual macroscopic review of actigraphy data to identify periods of daily activity less than or equal to that during confirmed night sleep periods. Evans et al[21] surveyed Old Order Amish by reporting the percentage of individuals that had napped in the previous week. Overall, subject reported sleep events have been found to be more accurate than actigraphy in determining discrete episodic events, particularly with respect to wake events.[7]

The ability of actigraphy to more accurately infer daytime sleep and nighttime wake would enhance the scope of studying sleep not only in mobile individuals within industrial societies, but also in subjects from nonindustrial societies in logistically challenging environments. One particularly salient issue in widely applying actigraphy to assess sleep, though, is that the specific parameter settings used to identify nap duration and activity counts can impact sleep-wake determinations. There are 2 levels at which software parameter settings can be adjusted to improve the detection rate of naps with actigraphy. The first is at the level of raw movement data, hereafter referred to as counts. In the process of implementing a sleep-scoring algorithm (such as that described in[22,] some threshold value is imposed on the data to distinguish likely-sleep (score of 0) from likely-wake (score of 1) on a per-epoch (ie, per-minute) basis. That binary data are then secondarily scored according to interval length of continuous 1 and 0 second. The maximum and minimum interval lengths used for this secondary scoring are also adjustable parameter settings. The first level is relatively device-specific, because the units of measure for raw-movement can vary by manufacturer, requiring that threshold values scale up or down accordingly. The second level can be readily compared across devices, because the binary data values are more biologically meaningful and once removed from the raw, device-specific data.

The present study aims to determine whether actigraphy can accurately detect episodes of reported daily nap and nighttime wake-bout events among a nomadic foraging population, the Hadza of Tanzania. We used the Motionwatch 8 (CamNtech) to collect actigraphy data and CamNtech's associated software to predict naps and wake bouts under different parameter settings. We aimed to identify the optimal parameters for detecting naps and wake bouts to test for the most reliable settings. Reliability was established in a 2-fold approach: rates of actigraphically identified naps were compared with participant self-report for agreement, and then the highest performing parameter configuration from this analysis was compared epoch-by-epoch to the method used in the previous study on Hadza sleep patterns,[12] where software settings were configured to identify 2 major sleep intervals per 24-hour period. Finally, we examined whether these settings have the capacity to discriminate the absence of napping.

Participants and methods

The Hadza are considered to be "median" hunter-gatherers, as they lie near the median value for most ecological and life history traits among subtropical hunter-gatherers.[23] They live in an east African environment with an effective temperature of 17°C, which is close to the median for warm-climate hunter-gatherers (16.3°C). The annual Hadza diet consists of approximately 43% hunted foods (game animals, birds, and honey) and 57% gathered foods [31] (median value for African hunter-gatherers is 32% and 67%, respectively), with high variability between seasons and across years. Their median local group home range is 122 km^2, although range sizes are declining due to adjacent population pressure and globalization.[24] The Hadza exhibit high levels of sexual division of labor, with males acting as primary hunters and females as primary gatherers. Importantly, the median local group size is approximately 30 for the Hadza (mean

group size for warm-climate hunter-gatherers is 26), and groups are characterized by central-place provisioning, with individuals returning to a central place to distribute food.[32] The most prominent material item used as a sleeping platform is a hardened impala skin and blankets. During some seasons, sleep is accompanied by the use of fire. Married couples sleep together with offspring, usually on the same bed. Seasonal variation in sleeping sites occurs; during the dry season, it is quite common to sleep outside, whereas during the rainy season, almost every individual sleeps within their thatched grass huts.

Participants were healthy adults older than 18 years, residing in nomadic or seminomadic camps that engaged in daily foraging. Thirty-three subjects completed the study, representing 21 women (mean age, 34.9 ± 14.3 years) and 12 men (mean age, 35.6 ± 14.7 years). Exclusion criteria included self-reported insomnia or physical disability due to injury or infirmed status that prevented an individual from engaging in active foraging. No individuals reported sleep problems, although 3 individuals were excluded from the analysis due to injury or infirmed status. All research was approved by the Tanzanian Commission for Science and Technology (COSTECH) and the Tanzanian National Institute for Medical Research (NIMR). All eligible subjects gave their verbal informed consent, as outlined by the institutional review board for human subjects research at the University of Nevada, Las Vegas, and Duke University.

We used the Motionwatch 8 actigraph (CamNtech), with all watches configured to generate data in 1-minute epochs. Subjects were asked to press the event marker preceding any sleep event throughout the study, including sleep after nighttime wake bouts and before initiating daytime naps. Participants were given a sleep survey at the beginning of the study period (January 20–February 11, 2016) to screen for healthy sleep, and they were instructed to not to remove watches throughout the study period. Participants also answered daily questionnaires (for every day throughout the study period) about the frequency of daytime sleep bouts (ie, napping) and discrete nighttime wake bouts. During daily questionnaires subjects were asked if they had pressed the event marker or if there were any technical difficulties, including instances of watch removal, during the previous day. There was not a single point throughout the study when a watch was observed to be removed, or reported to be removed by a participant. To ensure that nighttime wake bouts were discrete events requiring consecutive epochs of arousal between sleep onset and sleep end, we further identified nighttime wake bouts reported by nursing mothers.

Actigraph data were scored using the CamNtech MotionWare 1.1.15 program. The software has a nap analysis function that detects periods of inactivity that are attributable to napping or microsleeps. To achieve this, the software has parameters that the user can adjust. The first parameter of interest for this study, pertaining to the raw movement data itself, is hereafter referred to as the *nap activity threshold.* At this level, the epoch within a specified period must be less than or equal to the entered value for the period to be scored as a nap. For example, if the threshold is set to 10, the period will be scored as a nap if every epoch in a specified time period ≤10 activity counts (note: the units of an "activity count" and related actual bodily motion vary considerably from device to device, but can be adjusted in processing to yield comparable measures of sleep and activity across devices[25]). The second set of parameters pertain to the higher-level secondary processing of binary likely-sleep/likely-wake data, which we hereafter refer to as the *nap detection method*, because the software identifies nap episodes themselves. *Minimum nap length* sets the minimum number of epochs of inactivity required to be scored as a nap. *Maximum nap length* sets the maximum period of inactivity required to be scored as a nap. We compared sleep calculations derived from using these parameters to the automatic software scoring that relies on *a priori* knowledge of sleep segmentation to measures sleep, which we refer to as the *daytime sleep detection method*, because it does not identify nap episodes themselves. In applying this method, we identified examine sleep periods during the day, and then used the software's automatic sleep scoring to generate an epoch-by-epoch output classifying each epoch as either *sleep* or *wake*.

Using these different methods to create the sleep measures for the sake of comparison, we conducted 2 epoch-by-epoch (minute-by-minute) comparisons: (*i*) between reported daytime napping and actigraphy and (ii) between reported nighttime wake bouts and actigraphy. For daytime napping, we examined nine parameter setting combinations (Table 1), 3 pairs of min-max length intervals for napping (5-210, 15-210, and 40-210 minutes) each with 3 nap activity thresholds (10, 25, and 50). From our daily questionnaire, 137 days with at least one nap were reported among the 393 days of data collection. For comparative purposes, we randomly generated 30 days from the sample of 262 days in which no nap was reported by our subjects. Subjects wore the watches for an average of 14 days (range, 10-20).

For nighttime wake bouts, we generated raw count data and binary "sleep-wake" outputs from the MotionWatch software and imported them into R to identify consecutive periods of awake behavior during the nighttime sleep. We took a similar approach as the daytime nap analysis, but aimed to identify periods of wakefulness rather than periods of sleep in the actigraphy data. Specifically, we used the sleep-wake output and the raw count data and focused on 6 parameter settings in total (Table 2): for the sleep/wake output, we used 15, 20, and 30 consecutive wake epochs as determined by the MotionWare sleep algorithm. That is, a wake bout was counted if there were consecutive algorithm defined "wake" epochs. We also analyzed raw count data with 3 sets of parameters that were chosen on the basis that they showed no overestimation or underestimation bias for determining nap episodes—in other words, the most reliable settings from the daytime nap analysis (see results below). Comparisons between reported and actigraphic events were assessed for accuracy, sensitivity, and specificity.

Statistical analysis

Statistical analyses were conducted using *R* version 3.1.3.[26] Throughout, we pooled the actigraphy data of the participants for analysis. We used the Bland-Altman technique with the *BlandAltmanLeh* package to plot the difference between actigraphy and reported events against the average of actigraphy and reported events to determine whether there is a bias in actigraphy.[27] This technique was preferred because it is based on the agreement between 2 quantitative measurements,[28] and because it has been used in previous nap validation studies in controlled environments.[18] Plotting the

Table 1
The 9-nap detection parameter settings tested. Using these parameters, we assessed how well the actigraphy software detected reported events. The maximum duration for sleep during the day was set to 3.5 hours. Throughout the study, no naps longer than 3.5 hours were detected.

Parameter setting	Nap activity threshold (counts)	Min nap length (min)	Max nap length (min)
P1	10	5	210
P2	25	5	210
P3	50	5	210
P4	10	15	210
P5	25	15	210
P6	50	15	210
P7	10	40	210
P8	25	40	210
P9	50	40	210

Table 2
The 6 nighttime wake-bout detection parameter settings tested against reported events.

Parameter setting	Data type	Consecutive epochs with counts =>	Consecutive epochs with wake
Sleep-wake	Algorithm	N/A	15
Sleep-wake	Algorithm	N/A	20
Sleep-wake	Algorithm	N/A	30
P4 wake detection	Counts	10	15
P5 wake detection	Counts	25	15
P6 wake detection	Counts	50	15

We used the MotionWare output for sleep analysis that produces an epoch-by-epoch determination of *sleep* or *awake*. For comparison, we used the nap detection parameter setting (P6) that did not overestimate or underestimate naps and had the lowest mean difference; instead of assessing daytime sleep, however, they were used to assess nighttime wake events.

differences between 2 methods permits comparison of the mean of the measurements from both methods and reveals the distance (in the measurement units) of the gap between the *x*-axis and corresponding zero difference; thus, the distance of this gap illustrates the bias between the methods. To determine the significance of the mean difference obtained with the Bland-Altman technique, we used the upper and lower limits from the 95% confidence interval. The bias in actigraphy was represented as the mean difference between actigraphy and reported events, with positive mean difference indicating overestimation and negative mean difference indicating underestimation. We selected an *a priori* range of 0-0.5 naps as the biologically relevant threshold for measurement error.[28] This indicates that average deviations between tested methods of half a nap show that one method identifies a particular reported event that the other does not.

A day-by-day analysis was conducted for 167 total nights to determine both (i) agreement and (ii) accuracy, sensitivity, and specificity between 2 nap detection methods. In previous work with field actigraphy,[12] the daytime sleep detection method was used to analyze the daytime period to determine whether napping occurred during the day by applying the software's sleep algorithm to assess sleep during the day. A conservative, time-intensive method foregoes automated detection to visually assess each 24-hour output in the analysis by manually interfacing with the software to select each nap according to event markers and expert judging. Using this conservative method, we compared (i) the nap detection method that used the parameters with the greatest reliability found in this study (parameter setting 6, see results) against the (ii) daytime sleep detection method.

We used the definitions established by Tilmanne et al[29] to calculate accuracy, sensitivity, and specificity. When compared with self-reported events, a true positive (TP) indicates that the actigraph identifies napping correctly, a true negative (TN) indicates that the actigraph identifies the absence of napping correctly, a false positive (FP) indicates the actigraph misidentifies napping, and a false negative (FN) indicates that the actigraph misidentifies the absence of napping. Accuracy was defined as (TP + TN)/(TP + TN + FN + FP) and represents the agreement rate between manually scored self-reports and the actigraph. Sensitivity was defined as TP/(TP + FN) and represents the percentage of days identified correctly as having a nap. Finally, specificity is defined as TN/(TN + FP) and represents the percentage of days identified correctly as having an absence of napping. Because the data showed a nonnormal distribution, we used a Wilcoxon signed rank test to compare the total duration of daytime sleep (ie, that sum of minutes scored as nap behavior during the outside of bed daily period) estimated by both methods.

Results

Daytime sleep detection (napping)

A total of 167 days were analyzed. Proportionally, awake event markers were used 0.73 (SE = 0.02) of the total wake events and sleep event markers were used for 0.88 (SE = 0.03) of the time total sleep events. Results are presented in Table 3, with parameter settings labeled P1 to P9 and corresponding to sets of 3 parameter values given in Table 1 for minimum, maximum, and threshold. The Bland-Altman technique demonstrated that P6 (min-max of 15-210 minutes, with threshold ≤50 counts) shared the greatest agreement with reported napping events. Settings for P4, P5, and P6 produced results that were within the 0.5 *a priori* determination of successful nap detection, but P6 had the lowest mean difference (0.16). Based on the confidence intervals, P1, P2, and P3 settings overestimated daytime napping and P7, P8, and P9 settings for the 3 parameters underestimated daytime napping. P4, P5, and P6 neither overestimated nor underestimated napping significantly. Only settings in P6 did not differ significantly from reported events (see Fig. 1).

Nighttime wake-bout detection

Results are presented in Table 4. Fifty-five nights in which nursing mothers reported a presence or absence of wake-events were analyzed. The Bland-Altman technique demonstrated that the 20-minute sleep-wake algorithm, where 20 consecutive epochs are required to constitute a nighttime wake bout, is the method with the highest agreement with reported events (mean difference, 0.16). In addition, of the 3 sets of parameters that only analyzed raw count data and not data output from the sleep software's binary sleep-wake determination, the setting that had a threshold of ≥10 counts for 15 consecutive epochs (mean difference, 0.42) fell under the 0.5 *a priori* determination of successful wake-bout detection. Based on the confidence intervals, the 15-minute sleep-wake algorithm overestimated wake bouts and the 30-minute sleep-wake algorithm and other parameter sets underestimated wake bouts (Fig. 2).

Table 3
Bland-Altman statistics for tested agreement of nap detection parameter settings and reported nap events.

	P1	P2	P3	P4	P5	P6	P7	P8	P9
Mean difference	-1.56	−2.46	−4.21	0.37	0.30	0.16	0.83	0.81	0.75
Upper limit	3.30	3.60	4.25	2.17	2.11	2.09	2.14	2.11	2.08
Lower limit	−6.42	−8.51	−12.68	−1.42	−1.51	−1.77	−0.46	−0.48	−0.60
Critical difference	4.86	6.05	8.46	1.80	1.81	1.93	1.30	1.30	1.96
Estimation	Over	Over	Over	–	–	–	Under	Under	Under
Significance	*	*	*	*	*		*	*	*

The mean difference (estimated bias of actigraphy nap detection), upper and lower limits (95% confidence interval), and critical difference (2 times the SD of differences) are shown for all settings. Estimation indicates whether the parameter settings significantly overestimate or underestimate naps, compared with reported naps. A failure to reach the significance threshold indicates that actigraphy does not systematically mis-score the presence or absence of a nap; this was true only for P6.

Accuracy, sensitivity, and specificity

Compared with the "conservative" method, where event markers and self-reported naps are manually identified, the P6 nap detection method has the greatest accuracy and specificity, whereas the daytime sleep detection method has the greatest sensitivity (Table 5). In addition, a Wilcoxon signed rank test revealed that the daytime sleep detection method (mean, 38.8 ± 55.8) estimated greater total napping minutes when compared with manual scoring (mean, 27.2 ± 38.1; $P < .001$). We found no significant differences between the manual scoring and the results from using P6 settings (mean, 20.2 ± 35.8; $P = .19$), indicating that the P6 nap detection method better estimated the duration of time spent in daytime sleeping. In addition, when testing correspondence between the daytime sleep detection method and manual scoring, the Bland-Altman test revealed a mean difference in time spent daytime sleeping of −12.73 (upper limit = 59.7, lower limit = −34.1, critical difference = 46.9), with a tendency for overestimation. In contrast, when comparing total time spent daytime sleeping, the mean difference between the P6 nap detection method and manual scoring technique was 4.13 (upper limit = 40.4, lower limit = −31.1, critical difference = 36.3), with neither a tendency for overestimation or underestimation.

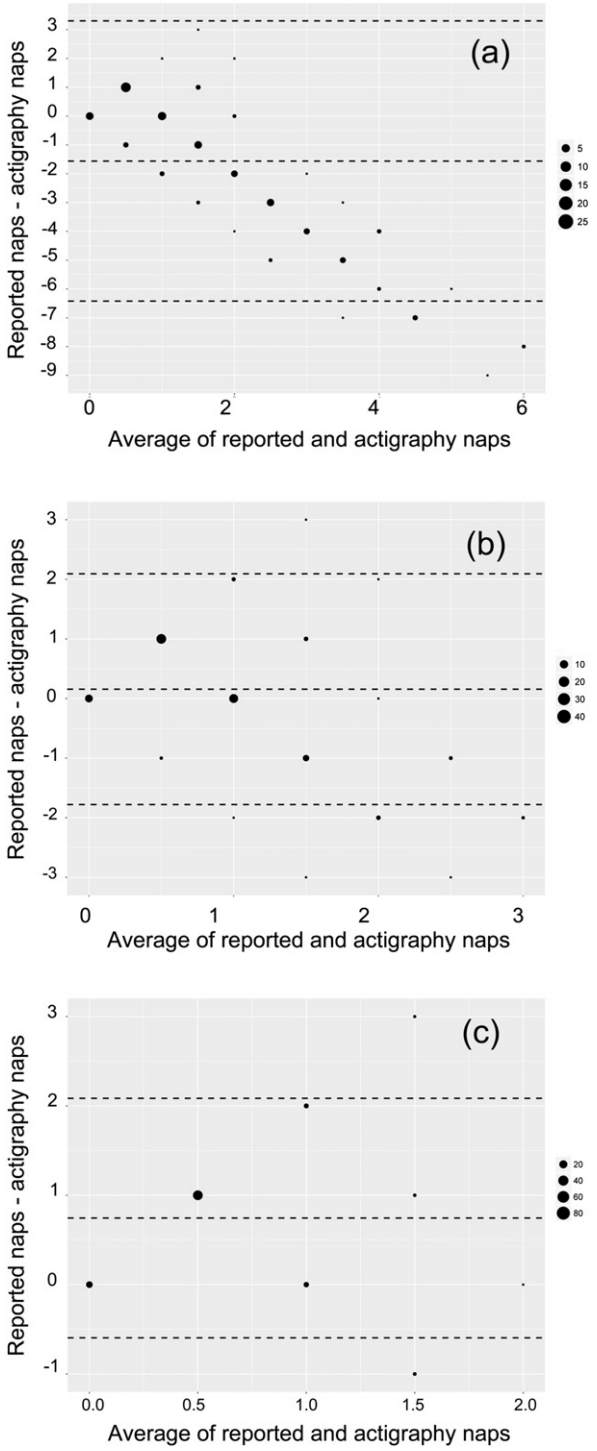

Fig. 1. Bland–Altman graphs for daytime napping. The graphs showed the P6 setting to be characterized by the least bias. Plotted difference between reported and actigraphic naps (*y*-axis), against the mean of both reported and actigraphy naps (*x*-axis; see middle dashed line), with 95% limits of agreement (the bottom and top dashed lines); the scale of dots indicates the number of overlapping data points (as indicated in the legend). Bias is indicated by the middle dashed line, where departures from zero mean difference (*y*-axis) indicate greater bias. A representative graph is displayed for each of the 3 min-max nap length settings: P1 (a), P6 (b), and P8 (c).

Discussion

This study examined the performance of an actigraphy-scoring algorithm to distinguish daytime sleep and nighttime wake bouts based on different parameter settings. The results suggest that actigraphy can be a useful tool for detecting sleep segmentation, yet calculated sleep measures are highly dependent on parameter settings. With respect to napping, the duration parameter, such as short (5 minutes) or long (40 minutes), can overestimate or underestimate napping, respectively. Parameter settings that captured moderate duration of inactivity (15 estimate) and large count interval (≤50) showed the greatest agreement with reported napping events; it was the only combination of settings that showed a nonsignificant difference with reported events and had the least overall mean difference. With respect to wake bouts, certain parameter settings, such as the sleep-wake detection algorithm coded to detect 15 consecutive minutes of wake during the nighttime sleep period, overestimated wake bouts, whereas the sleep-wake detection algorithm coded to detect 30 consecutive minutes of wake during the nighttime sleep period underestimated wake bouts. In addition, raw counts coded to detect 15 consecutive minutes of ≤25, or ≤50 counts underestimated wake bouts compared with reported events. The setting where the sleep-wake algorithm was parameterized to detect 20 consecutive minutes throughout the designated sleep period did not overestimate or underestimate wake bouts, had the lowest mean difference, and did not significantly differ from reported wake-bout events; thus, this method was the most reliable.

With respect to detecting daytime sleep bouts, the comparison between the daytime sleep detection method (where the sleep analysis algorithm is specifically applied to a segment of time *a priori* known to have a nap) and the nap detection method (which identifies nap events without *a priori* knowledge, using parameter settings P6; Table 1) showed support for the latter method as the most reliable. First, the P6 parameter setting showed the greatest accuracy (77%) for detecting both the presence and absence of naps, with notably high specificity in detecting the absence of naps (84%). In addition, when nap durations were compared between manual scoring and each of the 2 methods, the P6 parameter settings showed no difference in calculated total sleep durations. Thus, we recommend it as a useful approach for future detection of daytime sleep in traditional populations.

Based on this current analysis, we propose operational definitions for the 2 dimensions of sleep segmentation involving naps or nighttime awakenings in the context of actigraphy. Thorpy[20] defined a nap as any short sleep episode outside the bed without criteria on duration. Kanady et al[18] validated accelerometric measures of naps

against PSG, but using a method that required *a priori* knowledge of a napping event's existence. In trying to develop an algorithmic way of automatically identifying naps reliably (relative to self-report), we operationalized this definition within the context of actigraphy: a nap is constituted as a 15-minute period from the beginning of sleep end until sleep onset with an activity count of less than or equal to 50. In addition, for reports of wake bouts, we used the participants among the Hadza who were nursing mothers as benchmarks for biologically-relevant, sustained nighttime activity performed after sleep onset and before sleep end. Therefore, we operationalized this definition of segmented sleep as: a wake bout is constituted as a period of 20 consecutive minutes categorized as "awake" from the beginning time of sleep onset until sleep end. These operationalized definitions, pertaining to the secondary sleep scoring processes applied to binary likely-sleep/likely-wake values, are readily applicable for direct implementation of sleep scoring in any software package that allows for adjusting these parameters and should not be sensitive to any issues regarding interdevice differences, at least among the most commonly used and already validated sleep-monitoring accelerometer devices.

The use of self-reported data has several limitations. As discussed previously, PSG is the common "gold standard" for measuring sleep, yet its application in ambulatory subjects in remote, undeveloped areas presents several challenges to sleep researchers in field environments. Thus, we instead used self-reported instances of daytime napping and nighttime waking events to assess the performance of different settings. Self-reporting, however, can be problematic because it requires subjective recall and reporting of all events. Previous work in a sample of depressive insomniacs illustrated that self-reported sleep measures differed from PSG measures, although the pairwise correlation between PSG and sleep diary reports was significant.[30] In contrast, in healthy, noninsomniac subjects, Kawada[7] showed that sleep diary reports of daytime data were more valid for detecting sleep-wake activity than accelerometer data. In addition, our subjects were reminded each day, throughout the study duration, to press the event marker, which our analysis showed they did 88% of the time. This provided an added level of reliability to the reported event, given that most of the events were associated with a mark during the time of the event and not reliant on recall. Nevertheless, though, shorter naps, microsleeps, and shorter wake bouts may be less likely to be reported. Although likely meaningful in the context of physiological function and sleep regulation, these types of events are not generally considered the same type of behaviorally significant events discussed in the context of identifying sleep segmentation patterns. Overall, we recommend self-reporting as a way forward in validating actigraphy in field environments until technological innovation allows for field-based noninvasive PSG sleep staging.

Future work extending from this research could use portable PSG that have the capacity to measure sleep stages via electroencephalogram, electrooculogram, and electromyogram. Although challenging due to the limitations of battery life in nonelectric environments and the limiting aspects (eg, inducing low activity levels) of electrodes on participants throughout 24-hour use, the application of PSG would circumvent methodological issues that may arise from reporter bias.

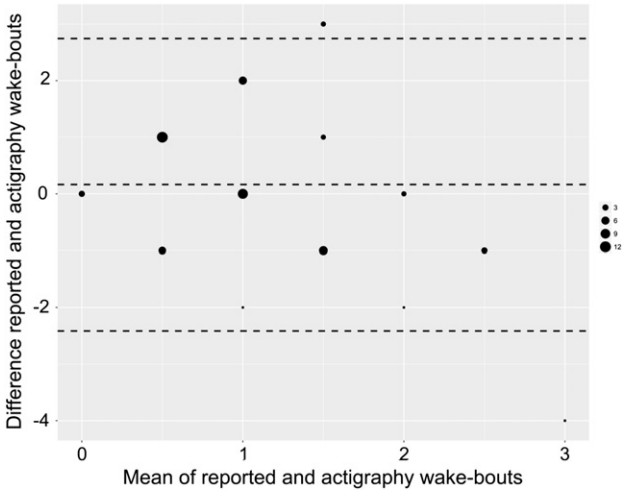

Fig. 2. The 20 minute *sleep–wake* algorithm, where 20 consecutive epochs are required to constitute a nighttime wake bout, is the method with the highest agreement with reported events.

Conclusion

Overall, this study suggests that actigraphy, with the appropriately designated parameters for nap and wake-bout diagnostics, can be used for detecting segmented sleep. Furthermore, Kanady et al[18] found that the high-sensitivity setting (ie, the threshold used to algorithmically determine sleep-wake states) used in most studies is optimal when investigators are interested in identifying all sleep epochs, but there is a bias of overscoring sleep. This has implications for future studies that aim to investigate segmented sleep, as such settings may bias against the identification of nighttime wake bouts, and studies may benefit from generating data form both low and high settings. The benefits of moving forward with an operationalized definition of the multiple dimensions of segmented sleep are great. Only with such definitions will future work be able to test predictions stemming from hypotheses directed at discovering the natural human sleep-wake pattern.

Disclosure

All authors state no conflict of interest.

Table 4
Bland-Altman statistics for tested agreement of nighttime wake-bout detection parameter settings and reported wake-bout events.

	S/W 15 min	S/W 20 min	S/W 30 min	P4 inverse	P5 inverse	P6 inverse
Mean difference	−0.69	0.16	0.81	0.42	0.8	1.02
Upper limit	2.43	2.74	2.51	3.52	2.81	2.72
Lower limit	−3.82	−2.41	−0.87	−2.68	−1.21	−0.69
Critical difference	3.1	2.57	1.69	3.10	2.01	1.70
Estimation	Over	–	Under	–	Under	Under
Significance	*		*		*	*

The mean difference (estimated bias of actigraphy wake-bout detection), upper and lower limits (95% confidence interval), and critical difference (2 times the standard deviation of differences) are shown for all settings. Estimation indicates whether the settings significantly overestimate or underestimate wake bouts at night, compared with the reported presence of wake bouts. A failure to reach the significance threshold indicates that actigraphy does not systematically mis-score the presence or absence of a wake bout.

Table 5
Accuracy, sensitivity, and specificity values compared 2 nap detection techniques.

	Sleep-during-day algorithm	Validated P6
Accuracy	66%	77%
Sensitivity	89%	68%
Specificity	47%	84%

Sensitivity indicates the ability of the test to detect napping (true-positive vs false-positive). Specificity measures how well the test detects the absence of napping (true-negative vs false-negative), whereas accuracy measures how well the test predicts both categories. Validated P6 from this study has the greatest accuracy and specificity, but not sensitivity.

Acknowledgments

We would like to thank the Hadza for their continued patience and the Tanzanian Commission for Science and Technology and the Tanzanian National Institute for Medical Research for research approval. In addition, we are grateful to the editor and the 2 anonymous reviewers for comments that added substantially to the content of this manuscript. This work was funded by National Geographic (Grant No. 9665-15).

References

1. Calogiuri G, Weydahl A, Carandente F. Methodological issues for studying the rest-activity cycle and sleep disturbances: a chronobiological approach using actigraphy data. *Biol Res Nurs*. 2013;15:5–12.
2. Cioffarelli A, Natale MT. Guide to the Catacombs of Rome and Its Surroundings[Bonsignori, Rome] ; 2000.
3. Rock P, Goodwin G, Harmer C, Wulff K. Daily rest-activity patterns in the bipolar phenotype: a controlled actigraphy study. *Chronobiol Int*. 2014;31: 290–296.
4. Wang J, Xian H, Licis A, et al. Measuring the impact of apnea and obesity on circadian activity patterns using functional linear modeling of actigraphy data. *J Circ Rhythm*. 2011;9:1-10.
5. Wright CE, Valdimarsdottir HB, Erblich J, Bovbjerg DH. Poor sleep the night before an experimental stress task is associated with reduced cortisol reactivity in healthy women. *Biol Psychol*. 2007;74:319–327.
6. Johnson NL, Kirchner HL, Rosen CL, et al. Sleep estimation using wrist actigraphy in adolescents with and without sleep disordered breathing: a comparison of three data modes. *Sleep*. 2007;30:899–905.
7. Kawada T. Agreement rates for sleep–wake judgments obtained via accelerometer and sleep diary: a comparison. *Behav Res Methods*. 2008;40:1026–1029.
8. Yoon IY, Kripke DF, Youngstedt SD, Elliott JA. Actigraphy suggests age-related differences in napping and nocturnal sleep. *J Sleep Res*. 2003;12:87–93.
9. Knutson KL. Sleep duration, quality, and timing and their associations with age in a community without electricity in Haiti. *Am J Hum Biol*. 2014;26:80–86.
10. Samson DR, Manus M, Nunn CL. Sleep duration, quality and timing in a non-electric population in Madagascar. *Am J Phys Anthropol*, conference preceedings. *Atlanta*. 2016.
11. de la Iglesia H, Fernández-Duque E, Golombek DA, et al. Access to electric light is associated with shorter sleep duration in a traditionally hunter-gatherer community. *J Biol Rhythms*. 2015;30:342–350.
12. Yetish G, Kaplan H, Gurven M, et al. Natural sleep and its seasonal variations in three pre-industrial societies. *Curr Biol*. 2015;25:1–7.
13. Calogiuri G, Weydahl A. Effects of sleep loss on the rest-activity circadian rhythm of helpers participating in continuous dogsled races. *Biol Res Nurs*. 2013;0:1–11.
14. Paul MA, Love RJ, Hawton A, Arendt J. Sleep and the endogenous melatonin rhythm of high arctic residents during the summer and winter. *Physiol Behav*. 2015;141:199–206.
15. Quante M, Kaplan ER, Rueschman M, et al. Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep. *Sleep Health*. 2015;1:275–284.
16. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*. 2003;26: 342–392.
17. Morgenthaler TI, Kapur VK, Brown T, et al. Practice parameters for the clinical evaluation and treatment of circadian rhythm sleep disorders. *Sleep*. 2007;30: 1705–1711.
18. Kanady JC, Drummond S, Mednick SC. Actigraphic assessment of a polysomnographic-recorded nap: a validation study. *J Sleep Res*. 2011;20: 214–222.
19. Carskadon MA, Dement WC. Normal human sleep: an overview. In: Kryger MH, Roth T, Dement WC, editors. Principles and Practice in Sleep Medicine. fourth edition. Philadelphia, PA: Elsevier Saunders; 2005.
20. Thorpy MJ. The international classification of sleep disorders: diagnostic and coding manual. *Am Sleep Dis Assoc*. 1990.
21. Evans DS, Snitker S, Wu SH, et al. Habitual sleep-wake patterns in the old order Amish: heritability and association with non-genetic factors. *Sleep*. 2001;34: 661–669.
22. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. *Sleep*. 1992;15:461–469.
23. Marlowe F. The Hadza: Hunter-Gatherers of Tanzania. Berkeley: University of California Press; 2010.
24. Crittenden AN, McKenna J, Gray P. Ancestral attachment: how the evolutionary foundation of attachment informs our understanding of child maltreatment interversions. In: Narvaez D, Valentino K, Fuentes A, editors. Ancestral Landscapes in Human Evolution: Culture, Childrearing and Social Wellbeing. Oxford University Press; 2014.
25. Cellini N, Buman MP, McDevitt EA, Ricker AA, Mednick SC. Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults. *Chronobiol Int*. 2013;30:691–698.
26. Team, R. C.. R: A Language and Environment for Statistical Computing[Vienna, Austria] ; 2015.
27. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 1995;346: 1085–1087.
28. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015; 25:141–151.
29. Tilmanne J, Urbain J, Kothare MV, Wouwer AV, Kothare SV. Algorithms for sleep-wake identification using actigraphy: a comparative study and new results. *J Sleep Res*. 2009;18:85–98.
30. McCall C, McCall WV. Comparison of actigraphy with polysomnography and sleep logs in depressed insomniacs. *J Sleep Res*. 2012;21:122–127.
31. Crittenden AN. Ethnobotany in evolutionary perspective: wild plants in diet composition and daily use among Hadza hunter-gatherers. In: Hardy K, Kubiak ML, editors. Wild harvest: plants in the hominin and pre-agrarian human worlds. Oxford: Oxbow Books; 2016. p. 319–339.
32. Marlowe FW. Central place provisioning, the Hadza as an example. In: Hohmann G, Robbins MM, editors. Feeding ecology in apes and other primates; 2006. p. 523.